Projecting impacts of uncertain climate change on future energy demand

Projecting impacts of uncertain climate change on future energy demand

Sandhya Patidar[1*], David P. Jenkins[1], Andrew Peacock[1]
[1]School of Energy, Geoscience, Infrastructure and Society, Heriot-Watt University, Edinburgh, UK
*Corresponding author's email address: S.Patidar@hw.ac.uk

Abstract

This paper presents a novel approach for developing a Demand Profile Synthesis Tool (DPSTool) involving the decomposition of electricity demand into subcomponents and selectively applying cutting-edge statistical modelling techniques for predicting (short/long-term) domestic energy demand patterns. The DPSTool will be underpinned by a new 'climate module' integrated within a previously developed high-tech HMM_GP model. The 'climate module' can simulate the impacts of key climatic variables on the electricity demand scenarios. The paper will demonstrate the key stages of statistical methodology development and validation procedure of 'climate module' using a residential case-study building selected from a community 'Auroville', located in India.

Key Innovations

- Development of 'Climate module' integrated within the core structure of HMM_GP modelling framework, detailed later in the paper, to generate long-term forecasts with high accuracies.
- DPSTool - A robust, rigorous and efficient system that integrates high-tech statistical and mathematical modelling techniques for simulating demand profiles at both individual and aggregated scale.

Practical Implications

The DPSTool will benefit, academic researchers and industry professionals working directly and/or in cross-sectional areas of data science (statistical/computational), energy demand modelling, building research/modelling community, energy systems/networks. The outcomes of the paper can be utilised by regulators, legislators, statutory consultees, government bodies including policymakers and implementers to understand the impact of future climate-related uncertainty on the energy system.

Introduction

Achieving an optimal balance of electricity demand with supply in real-time is one of the key challenges. Considering the future potentials for large-scale integration of renewables in the energy system, ability to predict electricity demand (in both the short and long term) with high accuracy is a central element of future strategic planning and policy development. Existing demand projection methods suffer from a range of limitations, including inadequate spatial/temporal resolution, reliance on rare/specific datasets (such as household diaries), and are not always amenable for application to future scenarios, often using just existing demand characteristics from historical data.

In this context, the availability of reasonably accurate energy demand projections/forecasts is important to ensure an optimally operating energy system which can support a range of applications. For example, optimising the generation process, stock pricing, controlling power systems, demand flexibility, resource allocation, planning and policy making (Torriti, 2020). New regulations require the energy industry to effectively integrate several environmental and regulatory factors in their practice, such as carbon emissions, volatility of the global energy market, optimised renewable energy usage, etc. To these several challenges, the ability to generate reliable energy demand patterns/forecast with a considerable amount of accuracy could lead to new insights towards future planning, resource allocation, infrastructure development, better customer services, increased potential for profit, reduced risk of failure, etc.. A considerable amount of novel research has been conducted recently, aligned towards an optimally operating energy system inclusive of generation, distribution and demand of energy (Centre for Energy Systems Integration (CESI) (Grant: EP/P001173/1), n.d.; Whole Systems Energy Modelling Consortium (WholeSEM: EP/K039326/1), n.d.; Adaptation and Resilience In Energy Systems (ARIES) (Grant:EP/I03534X/1, EP/I035773/1), n.d.). scenarios.

Energy system models, such as (MARKet ALlocation) (Taylor, Upham, McDowall, & Christopherson, 2014) and UKTM (the UK TIMES Model) (UKTM-UCL model website, n.d.; Daly & Fais, 2014), are conventionally used in the UK and across more than 70 countries by researchers and policy makers to understand the relationship between energy supply and demand over, usually, long periods of time (days/months) and across whole countries (or large regions of countries). These models can effectively simulate several future scenarios and thus can provide an understanding of cause and effect to assist policy development and designing future resilient energy systems. However, these energy system models are usually limited in terms of temporal and spatial resolution and quite often have constrained applicability when utilised for investigating the impact of factors influencing energy demand, i.e. they cannot be utilised to generate temporally precise energy demand profiles that can be varied with future change scenarios (e.g. climate change, policy change, new regulations, etc.). Specifically, with application to energy demand evolution in the domestic sector (Kannan, 2011), the highRES model (Zeyringer, Daly, Fais, Sharp, & Strachan, 2014; highRES model website, n.d.), utilising existing demand data (at hourly resolution) facilitates greater spatio-temporal detail for exploring potential impacts of future energy scenarios on the sustainability of whole energy systems. However, the highRES model only processes current demand data, making the quantification of issues such as future electrification of heat/transport, demand management/response and building-integrated storage more difficult to quantify.

When addressing such concerns, the availability of data-centric modelling approaches becomes more valuable. With significant advancement in technology (e.g. smart meters) our society is becoming more data rich than ever before, simultaneously creating opportunities to develop new data analytics approaches for providing better energy management services. Widely investigated statistical and hybrid approaches (that combine multiple modelling schemes to optimise predictive performance (Banihashemi, Ding, & Wang, 2017)) include: Box-Jenkins methods (e.g. Autoregressive Integrated Moving Average (ARIMA) Models (Erdogdu, 2007)), Artificial Neural Networks (ANN) (Kandil, Wamkeue, Saad, & Georges, 2006), Support Vector Machine (Dong, Cao, & Lee, 2005), Monte Carlo Simulation (Alani & Osunmakinde, 2017) and Markov Chain Models (Ullah, Ahmad, & DoHyeun, 2018). Most of these approaches are straightforward to implement, but are associated with certain limitations:

1) Complexity due to multiple causation factors – the functional relationship defining interactions of different factors (data types) influencing demand could be highly complex and thus overall complexity of any modelling schematic increases (while total reliability decreases) with the number of influencing factors included within the modelling structure;
2) Data Acquisition - model calibration usually requires detailed information on electricity and/or household diary datasets, and as such not always easily accessible;
3) Reliability of extremes - most of these techniques calibrated using observed records have limited reliability in the estimation of extreme (peak-demand) values (as not specifically designed to account for extreme values), and could be unsuitable for conducting impact analysis of unobserved extreme future scenarios;
4) Reliability at high spatial-temporal resolution.

Individual electricity demand profiles at high resolution (e.g. minutely) can communicate quite specific information about what is happening in a home and the impact that has on aggregated energy demand (whether at, for example, substation level or national level). The information can be statistically analysed to understand times of peak demand and, with other information available, the causes of those peak demands. However, if only a limited amount of metered/empirical data is available (and, at present, data at this resolution is still not widely available beyond a finite number of case studies), the availability of efficient statistical techniques, such as the "HMM-GP" model, become even more valuable (Patidar S. , Jenkins, Peacock, & Mccallum, 2019).

The HMM-GP model integrates Hidden Markov models (HMM) with a generalised Pareto Distribution (PD) through a robust STL (a Seasonal-Trend decomposition procedure based on Loess) time-series decomposition scheme (Cleveland, Cleveland, McRae, & Terpenning, 1990). Key details on the high-tech modelling structure of HMM-GP and its ability in simulating rigorous high-resolution electricity demand using observed historic records can be referred to in some preliminary work done by the authors (Patidar, Allen, Haynes, & Haynes, 2018; Patidar, Jenkins, & Simpson, Stochastic modelling techniques for generating synthetic energy demand profiles, 2016).

In this paper, we present our recent work utilising the potentials of multiple regression-based modelling approaches for synthesising climate perturbed energy demand profiles referred to as 'climate module'. The climate module is integrated within the complex architecture of the HMM-GP model and has been shown to simulated climate dataset available at a temporal resolution of 1 hour for simulating climate perturbed electricity demand profiles at 1-minute). The composite structure of the HMM-GP schematics that integrates novel climate module is referred to as 'DPSTool'. The DPSTool can be trained using observed historic records and has the potential to be applied at a large scale to simulate the impacts of future climate change (long/short term) on individual/aggregated demand profiles.

Data: organisation and analysis

To develop and demonstrate the development of the proposed 'climate module' for the HMM-GP model, the present paper will utilise energy demand dataset collected for a single case-study dwelling located in Auroville (Tamil-Nadu, India). Electricity demand data used for training and testing of the model are collected at a resolution of one minute for a period of 8 consecutive weeks in November - December 2018, starting from 1st November 2018 to 26th December 2018.

Selection of case-study for model development is mainly based on the criteria of availability of good-quality dataset, i.e. with less than 5% of missing points. The missing data are infilled using a logical approach developed by the authors and can be referred to elsewhere (Debnath, Jenkins, Patidar, & Peacock, 2020). The weather dataset for Auroville is available from the Meteoblue database (www.metoblue.com) at a temporal resolution of 1 hour for a period of 30 years (1989-2019). The key weather variable used in the development of the ‘climate module’ are: i) Temperature (T) measured at 2 m above the ground; ii) Relative Humidity (RH) measured at 2 m above the ground; iii) Total cloud cover (Tot_CC); iv) Sunshine duration (Sun_D); and v) Shortwave radiation (SR). The ‘climate module’ is calibrated using the first six weeks of the dataset and then rigorously tested using the remaining 2 weeks of the dataset. Key summary statistics of these five weather variables and electricity demand dataset over the eight weeks of the period are detailed in Table 1.

Method

In response to limited availability of high-resolution electricity demand data, the HMM-GP model, calibrated using a small sample of an available empirical dataset, has been extensively applied and demonstrated to generate synthesising “new” artificial demand profiles (Patidar S. , Jenkins, Peacock, & Mccallum, 2019; Patidar, Allen, Haynes, & Haynes, 2018; Patidar, Jenkins, & Simpson, Stochastic modelling techniques for generating synthetic energy demand profiles, 2016). One of the key features of the HMM-GP model is its ability to robustly generates synthetic demand series with the same statistical characteristics as the empirical data used for model calibration yet allowing these synthetic profiles to be sufficiently distinct to represent diversity. The modelling schematic has been intensively validated across a range of case studies. Moreover, these synthetic demand profiles have been shown to appropriately weighted and summated to replicate aggregated profiles of multiple dwellings, consistent with substation-level demand profiles, albeit with certain model limitations. Two key features responsible for the success of the HMM-GP modelling schematic are the integration of: i) STL-based time-series decomposition techniques; ii) Generalise Pareto (GP) distribution.

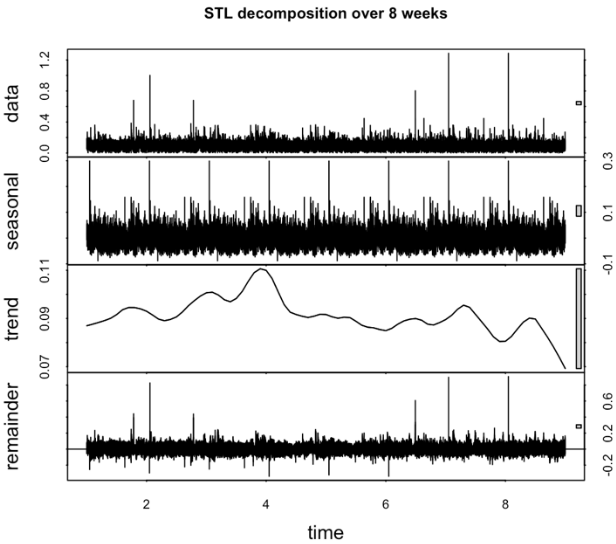

Figure 1 STL-based decomposition of observed electricity demand profiles (at one-minute resolution, kWh) of case-study dwelling over the 8 weeks periods (1st November 2018 – 26th December 2018).

The electricity demand profiles are comprised of various underlying endogenous (occupancy profiles, technology selection, etc.) and exogenous processes (climatic variability, seasonal effects, demographic variability etc.). These complex, nonlinear and partly stochastic time series integrate several superimposed, repeating activities occurring at different time periods (from hourly to annually). To simplify signals of this nature, the time series decomposition procedure deconstructs complex series by extracting deterministic components, i.e. “Trend” and “Seasonal” fluctuations, from the time series (Figure 1). The non-extracted, ‘residual’ series constitutes the “random” (stochastic) component of the time series. For STL decomposition, ‘seasonality’ should not be confused with meteorological seasons. In this paper, seasonality is defined at a level of weeks. The STL time series decomposition within the HMM-GP model facilitates a simplified modelling structure consisting of a single unit of HMM fitted to the stochastic component only, thus preserving deterministic features of the observed series in the simulation process, whereas integration of GP distribution (to values above 99th percentile) contributed towards effective modelling of extreme (peak demand) events. The peak demand values are of high importance as they constitute visible features in a typical electrical demand profile of a dwelling (e.g. kettles, cooking equipment, heating/cooling elements, electric showers). When aggregated with other dwellings, they contribute towards longer periods of energy use but due to the low frequency at which these features occur for an individual dwelling, data-driven models can struggle to simulate these values. The existing HMM-GP modelling procedure has been shown to stimulate these extreme peak demand values with considerably high accuracies.

The HMM-GP model, being entirely based on the observed dataset, assumes stationarity of future changes in the factors influencing electricity demand and thus is not suitable for impact analysis of future scenarios.

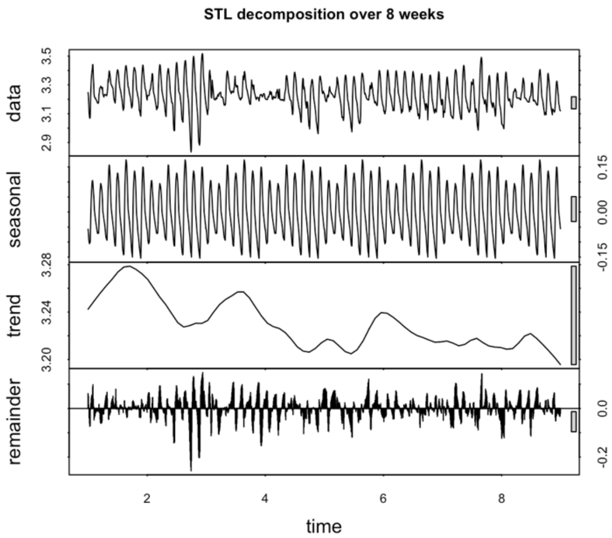

Figure 2 STL-based decomposition of log-transformed Temperature profiles (at one-hour resolution) over the 8 weeks periods (1st November 2018 – 26th December 2018).

The ***paper proposes novelty*** through the development of a robust modelling approach by reframing the HMM-GP modelling framework with a 'climate module'. The 'Climate module', is aimed to establish a statistical relationship for the deterministic components (i.e. trend and seasonality) of a few selected variable of weather variable (used as input variables) and the electricity demand profiles (output variables). Thus, the 'DPSTool' is a pioneering first step with the potential for allowing direct integration of climate change within the structural framework of the HMM-GP model. The "DPSTool' can be calibrated at a level of individual building demand profile. The calibrated model can be applied to a small sample of dwelling in a community to simulate synthetic profiles that can be scaled up to realise, and visualised, the impact in aggregated profiles of communities. This will allow the synthetic demand profiles to be future-morphed in a manner that reflects how individual dwelling profiles might respond to various future scenarios, and how this might be quantified at a regional, community level.

Calibrating 'Climate module'

The 'climate module' is calibrated using 6 weeks of the training dataset (1st November – 12th December 2018). Electricity demand data is identified as an output variable and five weather variables listed above (Table 1) is identified as an input variable. We aim to establish a simple statistical relationship using a multiple regression approach:

$$E = b_0 + b_1T + b_2RH + b_3Tot_{CC} + b_4Sun_D + b_5SR \quad \textbf{(1)}$$

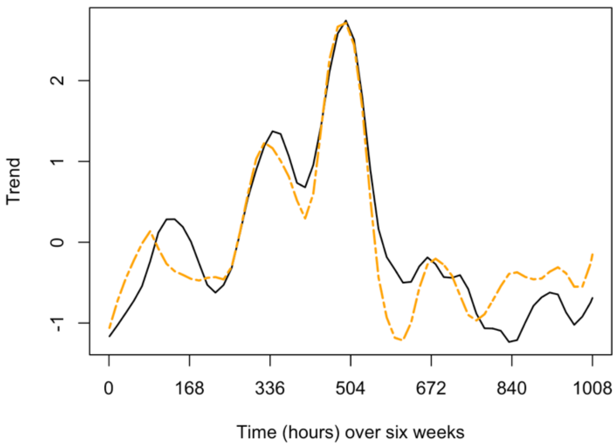

Figure 3 Comparing observed (in solid black line) versus predicted (in dashed orange lines) hourly trend component for electricity demand profiles over the six weeks (1st November – 12 December 2018) on the training dataset (1 week = 168 data points on x-axis).

The key steps involved in calibrating the 'climatic module' are:

1) **Scaling** - Estimating hourly averages of one-minutely electricity demand time series.
2) **Log-transformation**: Transforming multiplicative time-series into an additional time series by performing a log-transformation procedure of average electricity demand and all the five weather variables.
3) **STL-decomposition**: The hourly log-transformed time series of electricity demand and five weather variables are decomposed into "Trend", "Seasonal" and "Random" components through the application of STL-decomposition procedure, such that, for time series X, $X(t) = Trend(t) + Seasonal(t) + Random\,(t)$. STL-decomposition of log-transformed temperature profiles is presented in (Figure 2) for demonstration of STL application at hourly-scale.
4) **Standardisation** – The trend components of all the time series are standardised by subtracting their mean $(\bar{X})$ and diving standard deviation $\sigma(x)$, i.e. $X_{Trend}(t) = \frac{x(t) - \bar{X}}{\sigma(x)}$.
5) **Multiple Regression Model**: Fitting a multiple-regression model of Equation (1) type, to the standardised trend components. The regression model is fitted using the '*step*' function in R (Venables & Ripley, 2002). The results of regression analysis are presented in (Figure 3).

Analysis

This section aims to discuss the performance of a multiple regression model fitted to the trend components of average electricity demand with the five weather variables.

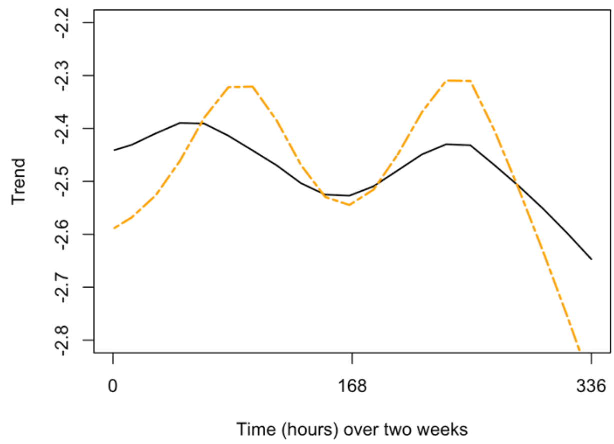

Figure 4 Comparing observed (in solid black line) versus predicted (in dashed orange lines) ***hourly*** *trend component for electricity demand profiles over the two weeks (13th December – 26 December 2018) on the test dataset.*

Assessing model performance on the training dataset

Table 2 displays the key summary statistics of the fitted model to assess the overall statistical significance of the model. **Call** displays the structure of the liner model, suggesting the optimal model involves all the five-weather variables. The 'step' function performs a thorough stepwise multiple regression analysis and identifies the optimum model using the AIC criterion. **Residuals** display the five summary statistics of the residuals/error values (Actual-predicted), which is approximately centred around 0.02 ~ 0.0 and the absolute value of the minimum and maximum are roughly around 0.84. Thus suggesting residuals are normally distributed with a centre around zero, indicating model assumption and inferences should be valid. Further, all the coefficients are statistically significant at (alpha-level 0.99, 99% confidence interval) in the model with a p-value much less than 0.05. A residual standard error that measuring the average distance between actual and predicted values is also considerably low ~ 0.38. Most important statistics, i.e. R-squared value is noted around 0.85. R-square is a goodness of fit measurement that tells us the proportion of variation in output variable that can be explained by multiple regression model. The R-squared value ranges between 0 to 1, and a value =1 indicates a perfect prediction of the output variable by the model. Thus, the observed R-square value of 0.85 can be considered high.

A considerably high value of F-statistics with an overall p-value of a model less than 0.05 ensure that the model is statistically significant. Finally, a considerably high correlation coefficient value of ~ 0.92 is noted for the observed versus predicted trend components of hourly average electricity demand.

To illustrate the performance of calibrated multiple-regression model on the training dataset, hourly values of observed trends (solid black line) are compared with the predicted trends (dashed orange lines) for electricity demand profiles for the entire duration of six weeks in (Figure 3). The figure shows a reasonably good match and the model appears to capture variation effectively.

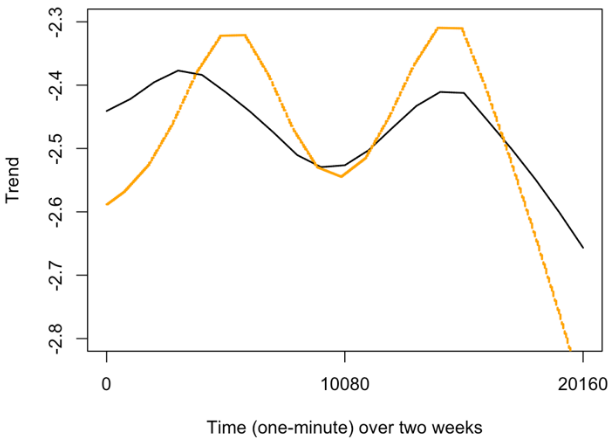

Figure 5 Comparing observed (in solid black line) versus predicted (in dashed orange lines) one-minutely trend component for electricity demand profiles over the two weeks (13th December – 26 December 2018) on the test dataset.

Assessing model performance on the test dataset

To assess the model's ability in projecting the 'trend' of unseen dataset a thorough model performance analysis is conducted. In Figure 4, hourly values of observed trends (solid black line) are compared with the predicted trends (dashed orange lines) for electricity demand profiles for the entire duration of two weeks of the test dataset in (Figure 4). A reasonable match, though weaker than the training dataset, is observed in the patterns of observed versus predicted profiles of hourly trend components. The slightly weak performance is naturally expected in the test dataset, but it is interesting to notice that the overall correlation coefficient measured is still considerably good 0.76. For the integrating 'climate module' within the DPSTool, the projected hourly average trend components are extrapolated to minutely scales. Figure 5 display the comparison of one-minutely values of observed (black solid lines) versus predicted (dashed orange lines) trend of electricity demand profiles for the entire duration of two weeks of the test dataset. Figure 5 is aimed to illustrate that the downscaling procedure applied at the climate dataset does not effect the dynamical characteristic of predicted energy demand at minutely scales.

Interestingly, the 'climate module' appears to perform reasonably well with a considerably small error (within a range of 0.2) and the extrapolation procedure appears to have no major impact.

Assessing model performance in simulating electricity demand within DPSTool (test dataset)

In this subsection, we assess the performance of the 'climate module' in predicting the electricity demand profiles for the test dataset (two weeks) using the HMM-GP model fitted to the training dataset (six weeks). For forecasting the electricity demand series over the entire period of two weeks of the test dataset we will utilise the electricity trend component predicted by the 'climate module' at an hourly scale, extrapolate it at one-minute resolution $\left(T_{predicted}\right)$. Since the seasonal components are extracted using a time frame of weekly length they do not vary over the weeks. Thus, we can extract the seasonal component for a complete 2 weeks periods from the HMM-GP model fitted to the training dataset $(S_{simulated})$. We will simulate the random component of the HMM-GP model (fitted to six weeks data) to forecast the random component of two weeks of the test dataset $(R_{simulated})$. We generate forecasted one-minutely electricity demand profiles for two weeks of the test dataset:

$$Forecast_{demand}(t) = T_{predicted}(t) + S_{Simulated}(t) + R_{Simulated}(t)$$

We compare key statistical properties of observed electricity demand profiles (Black solid line) with: i) synthetic demand profiles (Type 1 - Gray dashed lines) simulated using trend component extracted from the observed demand (using STL component of the HMM-GP schematic); and ii) with the synthetic demand profiles (Type 2 – Orange dotted line) generated using the trend component predicted by new 'climate module' integrated within DPSTool. Type 2 synthetic profiles can be generated merely using a climate dataset and an HMM-GP model fitted on a set of historic demand profiles. We use a consistent colour schematic for all the graphs compiled in Figures 6-7, presented to validate the efficiency of the 'climate module'.

Figure 6 presents an intensive comparison of i) Auto-correlation function (Left panel); ii) percentile distribution (Middle panel); and iii) probability density distribution (Right panel); for Observed (black solid lines) versus synthetic - Type 1 (Gray dashed lines) versus synthetic - Type 2 (Orange dotted lines) demand profiles. A considerably good match observed across all the key statistical properties confirming that 'climate module' can be utilised as a robust demand profiles synthesising tool.

Figure 7 illustrates the realisation of observed electricity demand compared with the synthetic (Type 1) and (Type 2) at one-minute resolution for a continuous period of one week. It can be visually seen that dynamics of observed matches reasonably well though synthetic demand (Type 2) appears to be slightly more spiker than (Type 1). Both the synthetic variants are more spiker than the observed profiles. The probabilities of high amplitude spikes are considerably low as can be seen in the probability density distribution plots with a secondary peak occurring at demand value 0.6 kW for synthetic Type 1 and 1.0 for synthetic Type 2 (Figure 6, right panel).

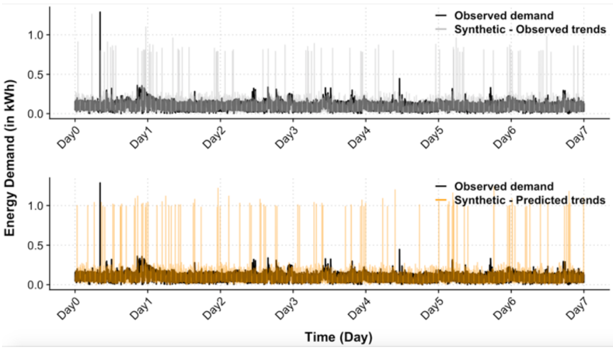

Figure 7: Comparison of the realisation over a week period, (Top panel) Observed electricity demand versus synthetic – Type 1 demands (Gray dashed lines) generated using observed trends; (Bottom panel) Observed electricity demand (black solid lines) versus synthetic – Type 2 (Orange dotted lines) demand profile.

Conclusion

The DPSTool presented herein is underpinned by novel synthetic time-series simulation techniques developed by the authors, referred to as HMM-GP. The HMM-GP is a suite of stochastic modelling techniques that integrate the Hidden Markov model (HMM) with the generalized Pareto Distribution (GP). The electricity demand decomposition element of the HMM-GP utilises a robust STL based time series decomposition approach for decomposing historical/observed electricity demand time series. The application of the STL approach facilitates the decomposition of complex dynamics (arising due to various superimposing deterministic (known) and stochastic (unknown) processes) of high-resolution electricity demand profiles into simple patterns referred to as components. Thus, allowing for a detailed analysis of these various underlying processes (known and unknown) constituting the demand in the simplified patterns of the trend, multiple seasonality, and random components.

The paper presents a statistical modelling procedure underpinning the 'climate module' that can be integrated within the HMM-GP model. The 'climate module' can be applied to simulate the impacts of climate change on demand. The method is developed and shown to predict demand at considerably high spatiotemporal resolution (one minute and at individual dwelling level). Thus, depending on the nature of the user-desired application, the proposed model can be scaled up and appropriately adapted to generate projections at low temporal resolutions (hourly, daily or even monthly). Previous work done by the authors has demonstrated the potential of the proposed approach in simulated community-level demand using the HMM-GP model using a small sample. The proposed 'climate module' integrated with the HMM-GP schematic has the potential to generate community-level future demand projection for a given climate projection with high accuracies ('accuracies' are depending on the accuracies of input climate projections).

The proposed approach could be a pioneering first step towards creating the possibility of direct integration of factors influencing demand profiles (i.e. contextual data) within the structural framework of the HMM-GP model. In the paper, we demonstrated that the decomposed time series of climatic dataset and electricity demand can be statistically analysed for identifying and establishing an underlying association of deterministic components (specifically in the trend features). A novel correlation module is developed and integrated within the framework of the HMM-GP model, referred to as the DPSTool, for conducting a thorough correlation analysis of the impacts of key climate variables on the energy demand. Thus, the DPSTool can be utilised to facilitate impact analysis of various future change scenarios, directly applied to individual profiles, while realising effect in aggregated profiles, i.e. at a regional, community level.

This paper mainly focused on demonstrating the capabilities of the DPSTool for a single dwelling and is limited to include the association in the trend features only, however, a similar analysis could be done to further explore the association in seasonal components. The DPSTool need to be extensively validated not just to ensure the robustness of the conceptual design of the modelling framework but to access its overall stability and performance across a range of dataset/case studies. The HMM-GP model has the potential to impact a range of applications as specified above and beyond, such as to understand energy use at a fine temporal scale, to generate aggregated demand profiles, and to understand the impact of various factors influencing energy demand usage (such as climate).

Acknowledgement

This work has been done as part of EPSRC funded project "Community-scale Energy Demand Reduction in India (CEDRI: EP/R008655/1).

References

Adaptation and Resilience In Energy Systems (ARIES) (Grant:EP/I03534X/1, EP/I035773/1). (n.d.). Retrieved 07 17, 2018, from http://www.arcc-network.org.uk/aries/

Alani, A. Y., & Osunmakinde, I. O. (2017). Short-Term Multiple Forecasting of Electric Energy Loads for Sustainable Demand Planning in Smart Grids for Smart Homes. *Sustainability, 9*, 1972.

Banihashemi, S., Ding, G., & Wang, J. (2017). Developing a Hybrid Model of Prediction and Classification Algorithms for Building Energy Consumption. *Energy Procedia, 110*, 371-376.

Centre for Energy Systems Integration (CESI) (Grant: EP/P001173/1). (n.d.). Retrieved 07 17, 2018, from https://www.ncl.ac.uk/cesi/

Cleveland, R. B., Cleveland, W. S., McRae, J. E., & Terpenning, I. (1990). STL: A Seasonal-Trend Decomposition Procedure Based on Loess. *Journal of Official Statistics, 6*(1), 3-73.

Daly, H. E., & Fais, B. (2014). *UK TIMES MODEL OVERVIEW.* London, UK: UCL Energy Institute. Retrieved from http://www.wholesem.ac.uk/documents/uktm-documentation

Debnath, K. B., Jenkins, D. P., Patidar, S., & Peacock, A. (2020). Understanding Residential Occupant Cooling Behaviour through Electricity Consumption in Warm-Humid Climate. *Buildings, 10*(4), 78.

Dokumentov, A., & Hyndman, R. J. (2015). STR: A Seasonal-Trend Decomposition Procedure Based on Regression. Department of Econometrics and Business Statistics, Monash University: Monash Econometrics and Business Statistics Working Papers 13/15.

Dong, B., Cao, C., & Lee, S. E. (2005). Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings, 37*(5), 545-553.

Erdogdu, E. (2007). Electricity demand analysis using cointegration and ARIMA modelling: a case study of Turkey. *Energy Policy, 35*(2), 1129–46.

highRES model website. (n.d.). Retrieved 07 05, 2018, from http://www.ucl.ac.uk/energy-models/models/highres

Jenkins, D. P., Patidar, S., & Simpson, S. A. (2014). Synthetic electricity demand profiles for UK dwellings. *Energy and Buildings, 76*, 605-614.

Kandil, N., Wamkeue, R., Saad, M., & Georges, S. (2006). An efficient approach for short term load forecasting using artificial neural network. *Electrical Power and Energy Systems, 28*, 525–30.

Kannan, R. (2011). The development and application of a temporal MARKAL energy system model using flexible time slicing. *Applied Energy, 88*(6), 2261-2272.

Patidar, S., Allen, D., Haynes, R., & Haynes, H. (2018). Stochastic modelling of flow sequences for improved prediction of fluvial flood hazards. *Geological Society, London, Special Publications, 488*, 205-219.

Patidar, S., Jenkins, D. P., & Simpson, S. A. (2014). Generating Synthetic Energy demand profile at one minute resolution – A statistical modelling approach. *Building simulation and Optimization (BSO14.ORG.* UCL, London, United Kingdom.

Patidar, S., Jenkins, D. P., & Simpson, S. A. (2016). Stochastic modelling techniques for generating synthetic energy demand profiles. *International Journal of Energy and Statistics, 4*(3), 1650014.

Patidar, S., Jenkins, D. P., Peacock, A., & Mccallum, P. (2019). Time Series Decomposition Approach for Simulating Electricity Demand Profile. *Building Simulation 2019, 16th IBPSA International International Conference.* Rome, Italy.

Taylor, P. G., Upham, P., McDowall, W., & Christopherson, D. (2014). Energy model, boundary object and societal lens: 35 years of the MARKAL model in the UK. *Energy Research and Social Science, 4*, 32-41.

Torriti, J. (2020). Household electricity demand, the intrinsic flexibility index and UK wholesale electricity market prices. *Environmental Economics and POlicy studies*.

UKTM-UCL model website. (n.d.). Retrieved 07 05, 2018, from http://www.ucl.ac.uk/energy-models/models/uk-times

Ullah, I., Ahmad, R., & DoHyeun, K. (2018). A Prediction Mechanism of Energy Consumption in Residential Buildings Using Hidden Markov Model. *energies, 11*, 358.

Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistic with S.* Springer.

Whole Systems Energy Modelling Consortium (WholeSEM: EP/K039326/1). (n.d.). Retrieved 07 17, 2018, from http://www.wholesem.ac.uk/about-the-project

Zeyringer, M., Daly, H., Fais, B., Sharp, E., & Strachan, N. (2014). Spatially and Temporally Explicit Energy System Modelling to Support the Transition to a Low Carbon Energy Infrastructure – Case Study for Wind Energy in the UK. *International Symposium for Next Generation Infrastructure Conference Proceedings.* IIASA, Laxenburg, Austria.

Table 1: Summary statistics of Electricity demand data at 1 minute and weather variables at 1 hour resolution.

Variables	Minimum	1st Quantile	Median	Mean	IIIrd Quantile	Maximum
Electricity demand (E) in kW	0.00	0.04	0.08	0.09	0.14	1.29
Temperature (T)	17.02	23.82	25.10	25.41	27.02	33.69

Relative Humidity (RH)	32.00	72.00	84.00	80.00	91.00	98.00
Total cloud cover (Tot_CC)	0.01	2.70	30.00	45.88	100.00	100.00
Sunshine duration (Sun_D)	0.01	0.01	0.01	15.49	40.85	60.00
Shortwave Radiation (SR)	0.01	0.01	2.67	183.24	346.21	865.08

Table 2: Outcomes of 'summary' function -Summary statistics of the fitted multiple regression model in R

Call:				
lm(formula = Trend_Avg_Electricity_demand ~ Trend_ Temp + Trend_ Rel_Hum + Trend_ Tot_CC + Trend_ Sun_D + Trend_SR)				
Residuals:				
Min	***1Q***	***Median***	***3Q***	***Max***
-0.84206	*-0.28878*	*-0.02617*	*0.27581*	*0.84536*
Coefficients:				
	Estimate	***Std. Error***	***t value***	***Pr(>|t|)***
(Intercept)	*-2.834e-16*	*1.214e-02*	*0.000*	*1*
Trend_ Temp	*4.757e-01*	*1.774e-02*	*26.819*	*< 2e-16 ****
Trend_ Rel_Hum	*-1.195e+00*	*2.788e-02*	*-42.876*	*< 2e-16 ****
Trend_ Tot_CC	*1.646e-01*	*3.176e-02*	*5.182*	*2.65e-07 ****
Trend_ Sun_D	*4.411e-01*	*4.688e-02*	*9.410*	*< 2e-16 ****
Trend_SR	*-2.035e+00*	*4.377e-02*	*-46.482*	*< 2e-16 ****
*Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*				
Residual standard error: 0.3856 on 1002 degrees of freedom				
Multiple R-squared: *0.8521*	*Adjusted R-squared: 0.8513*			
F-statistic: 1154 on 5 and 1002 DF	***p-value: < 2.2e-16***			
Correlation observed versus predicted (training data – 6 weeks)	*Correlation observed versus predicted (2 weeks test data)*			
0.9230791	*0.7685673*			

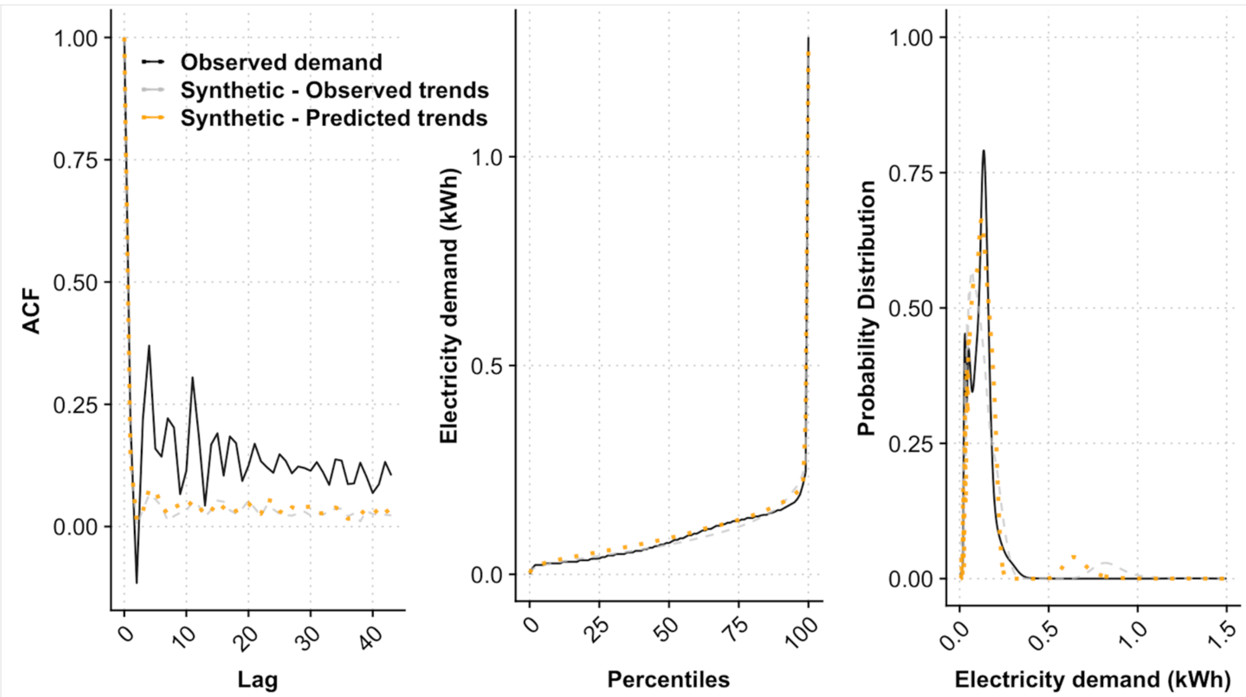

Figure 6: (Left panel) compares the Auto correlation function; (Middle panel) percentile distribution; (Right panel) probability density distribution; for Observed (black solid lines) versus synthetic - Type 1 (Gray dashed lines) versus synthetic - Type 2 (Orange dotted lines) demand profiles.